Pergamon

0042-6989(94)00209-6

Vision Res. Vol. 35, No. 10, pp. 1463–1472, 1995

0042-6989/95 $9.50 + 0.00

Interaction of Stereo and Texture Cues in the Perception of Three-dimensional Steps

JOHN P. FRISBY,*† DAVID BUCKLEY,* KEITH A. WISHART,* JOHN PORRILL,* JONAS GÅRDING,‡ JOHN E. W. MAYHEW*

Received 24 November 1993; in revised form 9 August 1994

A computational method for calibrating stereo using shape-from-texture is described together with five experiments that tested whether the human visual system implements the method. The experiments all tested the prediction that the perceived size of a step between two planar and slanted real surfaces should be affected by texture slant cues projected on to them that are inconsistent with the disparity cues. The predicted effect was observed but the results could be accounted for by a new phenomenon revealed in control conditions: the perceived size of a step between two slanted planes is in part determined by the size of the slants even when texture and stereo cues are held consistent. We conclude that the hypothesis that human stereo is calibrated by texture is not confirmed.

Stereopsis   Stereo calibration   Texture   Cue integration

INTRODUCTION

Consider the binocular viewing geometry viewing shown in Fig. 1. Starting from the analysis provided by Mayhew (1982), Porrill, Mayhew and Frisby (1985) provided an equation for the horizontal disparities generated by this geometry which showed that the horizontal disparity (H) for a given point is composed of four terms. These terms describe contributions due to retinal location [or eccentricity (ecc)], gaze angle (g), elevation angle (e) and the distance of a scene element from the plane containing the fixation point and lying perpendicular to the line connecting the fixation point to the "cyclopean eye" located midway between the eyes (z), as follows:

$$H = H_{ecc} + H_g + H_e + H_z$$

which produces the following equation

$$H = \frac{x^2 I}{d} + \frac{gxI}{d} + \frac{eyI}{2d} + \frac{zI}{d^2} \qquad (1)$$

where I is interocular separation and d is the distance to the fixation point.

This equation is an extension of that provided by Mayhew (1982) to cover the case where the eyes can be elevated out of the horizontal plane. The simplicity of the equation stems from the assumptions that I is small with respect to d and that all angles are small (<20 deg).

The important point about this equation for present purposes is that if metrical descriptions of scene structure are to be recovered from H (e.g. a distance z in cm), then the parameters d, g, and e, which vary as different points are fixated, need to be known, as well as I (which is of course constant in human vision). Obtaining these parameters is sometimes called the *stereo camera calibration problem* (or the *relative orientation problem* in photogrammetry). It is fundamental‖ to stereoscopic constancy: the metrical structure of the world appears to remain unchanged as the eyes roam over a stable scene despite dramatic changes to the retinal disparity field caused by eye movements.

How then are the required d, g, and e parameters to be recovered? There have been two main approaches to this problem and we propose here a third. Any given visual system might use any combination of these techniques as they are not mutually exclusive.

1. Use information from extra-retinal sources

This is the classic solution: obtain knowledge of eye positions from the oculomotor and/or accommodation

*AI Vision Research Unit, University of Sheffield, Sheffield S10 2TN, England.
†To whom all correspondence should be addressed [*Email* jpf@aivru.sheffield.ac.uk].
‡Royal Institute of Technology, Stockholm, Sweden.

‖Although this is true if metrical scene information is to be recovered, various qualitative attributes of scene surfaces, such as planar vs curved, can be obtained without explicit knowledge of d, g, and e. Gårding, Porrill, Mayhew and Frisby (1995) provide an extensive review of this question. They analyse the inter-relationships between the work of Koenderink and van Doorn (1976) on using the def component of the disparity gradient field, and Weinshall (1990) and Liu, Stevenson and Schor (1993) on using polar disparities. Gårding *et al.* also propose a new theory, showing how various qualitative surface atrributes can be obtained from a quantity called *scaled relative nearness* derived from horizontal disparities by using vertical disparities but in a way which does not make d, g, and e explicit.

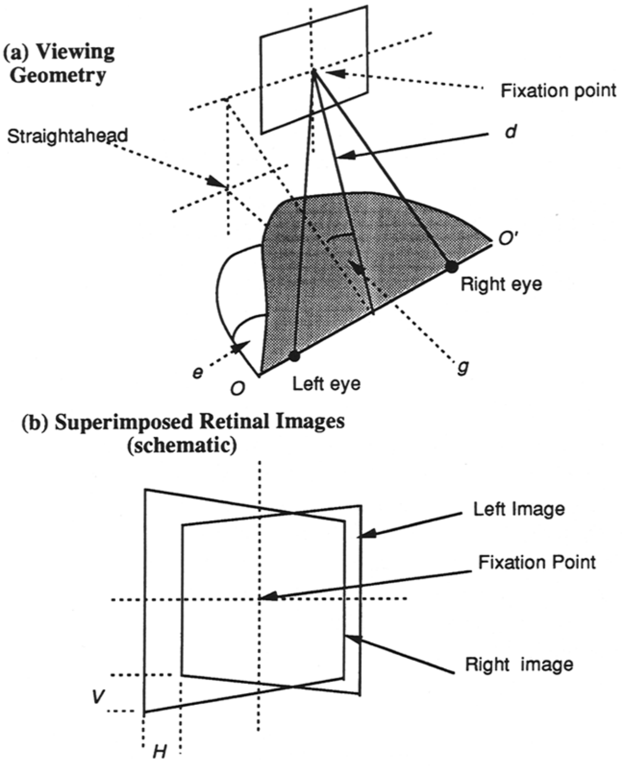

FIGURE 1. Binocular viewing geometry. (a) Two eyes are shown fixating on a point at distance d offset from straightahead by a gaze angle g and an elevation angle e. (b) Schematic representation of superimposed left and right retinal images showing horizontal (H) and vertical (V) disparities. Reprinted with permission from Frisby and Pollard (1991).

system, using either in-flow or out-flow signals. The viability of this approach has been questioned on the ground that judgements of depth from oculomotor/accommodation alone are notoriously poor (Foley, 1980, 1985; review in Collett, Schwarz & Sobel, 1991). This may relate to the high variability reported in measures of human vergence, which have led to doubts that it would be possible to obtain reliable parameterisation from keeping track of eye movement positions (Collewijn & Erkelens, 1990). On the other hand, Enright (1991) has recently reported psychophysical evidence for the exquisite control of binocular eye movements, and for highly precise depth judgements made from oculomotor information. His work fits in well with the computer vision studies of Mayhew, Zheng and Cornell (1993). They demonstrated the feasibility of neurophysiologically-plausible neural net techniques for achieving the accurate parameterization of a variable-geometry 4 degrees of freedom stereo camera rig. Their techniques depend only on the ability to fixate binocularly, not on measuring disparity. The errors in camera rig positions deriving from the resolution of the control motors were similar to equivalent data for human eye positions (Carpenter, 1988). This work therefore suggests that the human visual system could in principle use oculomotor information for stereo calibration but it remains a controversial question whether it actually does so.

2. Use vertical disparities

Mayhew (1982) provided not only a simple equation for H but also a remarkably simple equation for V, the vertical component of disparity vectors (see also Mayhew & Longuet-Higgins, 1982; Longuet-Higgins, 1982). Porrill, Mayhew and Frisby (1985) extended Mayhew's equations for V to take account of eye elevation angle, showing that V has three components (to first order)

$$V = V_{\text{ecc}} + V_g + V_e$$

which are given by

$$V = \frac{xyI}{d} + \frac{gyI}{d} + \frac{exI}{2d}. \tag{2}$$

The attractive property of this equation is that it shows V to be unaffected (to first order) by scene structure because there are no terms involving z. This opens up the prospect of solving for the unknown d, g, and e parameters using measures of V. This could in principle be done from just three points, but obviously more stable estimates might be achieved by pooling data from many points (Stenton, Frisby & Mayhew, 1984). Whether the human visual system uses vertical disparities in this way is currently a controversial issue (Frisby, 1984; Rogers, 1991, 1992). Cumming, Johnston and Parker (1991) and Sobel and Collett (1991) reported failure to find expected effects on perceived surface shape from scaling V but Rogers and Bradshaw (1993) have found the predicted effects for a large fields of view (see also Gårding *et al.*, 1995).

3. Use information from shape-from-texture

This approach was first noted, to our knowledge, in Porrill, Gårding, Eklundh, Frisby, Buckley, Pollard, Mayhew and Spivey (1991). The key idea is to use measures of surface slant from a shape-from-texture module to calibrate stereo. One method of this kind starts with Mayhew's (1982) observation that if it can be assumed that disparities derive from a surface that can be modelled as a set of locally planar patches, then the z term in his equation for H could be re-written as follows:

$$H = \frac{x^2I}{d} + \frac{gxI}{d} + \frac{eyI}{2d} + \frac{IPx}{d} + \frac{IQy}{d} + \frac{IK}{d^2}. \tag{3}$$

Suppose now that the P and Q parameters of a planar patch were to be provided by a shape-from-texture module, which is an entirely feasible proposal computationally (e.g. Gårding, 1992). P and Q information from texture could then be substituted into the H equation to solve for the global viewing parameters d, g, and e, plus the local offset K, using points arising from the planar patch. As few as four points would be needed in principle but again pooling techniques could be used if more data were available. The details of this scheme are set out in the Appendix.

A question immediately arising from this proposal is, if texture can already provide an estimate of P and Q, why bother to calibrate stereo with this information if the best that can then be expected from stereo is to recover the very

TABLE 1. Summary of experimental conditions

Experiment	N	Texture type	Step sizes (cm)	Stereo cue slants (deg)	Texture cue slants (deg)	Density cue	Viewing distance (cm)	Wall/ground plane
Experiment 1	6	Circles and dots	6, 8, 10, 12	0	10, 0, −10	Consistent with stereo step size	195 cm	Ground
Experiment 2	6	Circles	6, 8, 10, 12	10, 0, −10	10, 0, −10	Consistent with stereo step size	195 cm	Ground
Experiment 3	6	Dots	6, 8, 10, 12	10, 0, −10	10, 0, −10	Consistent with stereo step size	195 cm	Ground
Experiment 4	5	Squares	10, 15, 20	−10, −20, −30	−10, −20, −30	Consistent with stereo step size	150 cm	Wall
Experiment 5	10	Squares	7.5 to 22.5	−17, −20, −23	−17, −20, −23	Consistent with predicted step size	150 cm	Wall

same estimates of P and Q? One answer is that, having achieved a d, g, e calibration in this way, disparity information can then provide descriptions of scene structure other than just P and Q values. For example, the size of step discontinuities in the scene could be recovered. This would be valuable because texture is in general ill-suited to the recovery of metric information about steps. This is the consideration that underpinned the series of experiments reported here. Specifically, we tested whether judgements of the size of a step from stereo cues is altered when the step is set in texture fields carrying cues for a Q slant. This would, if texture were used to calibrate d, g, and e, lead to predictable alterations in step size perceptions (see Appendix for the method used to generate predictions). We have conducted a series of five experiments on this theme. The different features of their design are summarized in Table 1, but we begin by describing methodology common to them all.

METHODS

Task and apparatus

The observer's task was to judge the magnitude of a step set into either a ground plane [Fig. 2(a)] or a wall plane [Fig. 2(b)] viewed through a closely-fitting head-rest. This restricted the field of view to a roughly circular region of about 20–22 deg in diameter around the step, with the border of this region created by an aperture with a jagged outline. In Expts 1–3, viewing distance to the centre of the step was 195 cm, and stimulus durations were 4 sec; in Expts 4–6 these parameters were 150 cm and 5 sec respectively. Stimulus presentation times were controlled by a shutter attached to the head-rest. Disparity cues were carried by texture elements projected from a 35 mm Kodak carousel projector on to the ground or wall planes, with texture lines appearing as bright on a black ground. This ensured that the texture of the polystyrene surface was itself invisible to the observer. The projector slides were photographs of laser-printed texture fields generated by a computer graphics program which ensured that slant-from-texture cues could be made to be either consistent or inconsistent with the stereo cues. The slant-from-texture cues were successful, in the sense that introspective reports from observers as well as from ourselves indicated that the texture cues clearly altered the perceived slant of the planes forming the step. This is in keeping with the similar effects obtained by Frisby and Buckley (1992) who also used the technique of projecting textures carrying a slant cue on to real planar surfaces. The size of the slant-from-texture effect was not, however, measured in the present study.*

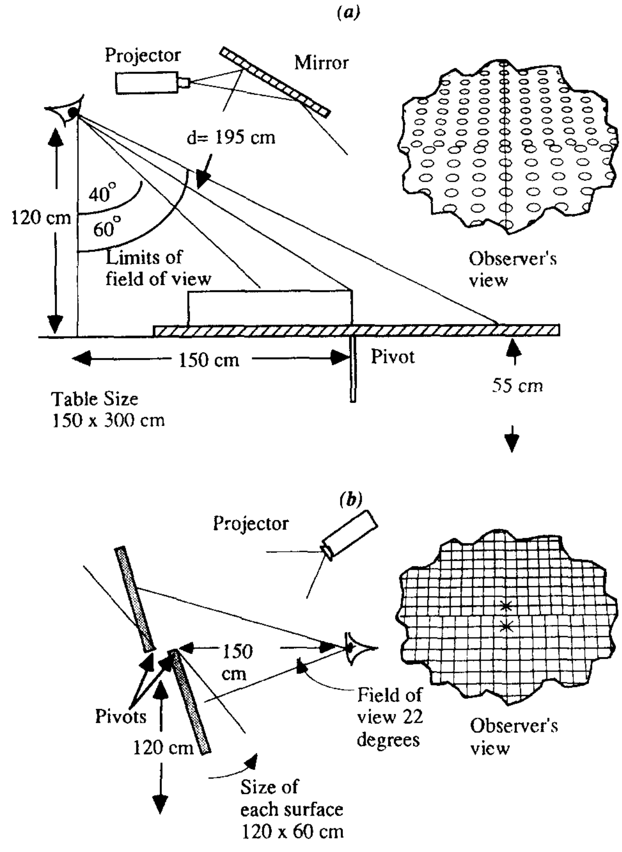

FIGURE 2. Apparatus. (a) Experiments 1–3, viewing distance 195 cm, diameter of field of view c. 20 deg. Observers were asked to judge the size of the step at the centre of the field of view marked with lines perpendicular to the step. (b) Experiments 4 and 5, viewing distance c. 150 cm, diameter of field of view c. 22 deg. The locii for step size judgements were marked by crosses. The intrinsic ambiguity of regular repeating patterns on planar surfaces allows them to be fused in different ways (the *wallpaper illusion*). The marker lines and crosses provided a check on this potential problem: observers were asked to report if they ever saw doubled marker lines or doubled crosses, indicating incorrect fusion, in which case the stimulus was shown again. This occurred very infrequently.

*A referee has raised the question of whether the well-known underestimation of visual texture slant by human observers (Turner, Gerstein & Bajcsy, 1991) could have had an important bearing on our results. Whereas that factor might have to be taken into account in a quantitative test of the idea that human vision might use texture to calibrate stereo, we think it can be neglected for the present experiments which sought qualitative evidence of the existence of texture slant effects on a step size judgement from stereo cues.

The main feature of the experimental set-up was that, by projecting the texture on to a real object (a step), it used "quasi-natural" stereo cues (Frisby & Buckley, 1992) (the qualification quasi- refers to the fact that the observer's head position was fixed and the field of view, although reasonably large, was not a full field). This way of creating stereo cues has the attractive feature that the mechanisms processing binocular disparity and those controlling accommodation and vergence are stimulated much as they would be for normal viewing of natural scenes. This is not the case in typical stereogram presentations for which the sykinesis between accommodation, disparity, and vergence is usually disrupted (Buckley & Frisby, 1993).

Observers

The 33 observers (22 men and 11 women, aged between 18 and 35 yr) were volunteers with normal or corrected-to-normal vision. The numbers taking part in each study are shown in the figures giving the data for each experiment. Most had not previously participated in a psychophysical experiment and each served for only one experiment. All observers were screened using the Titmus Randot Test (criterion for inclusion was stereoacuity of 50 sec arc or better).

Design

All experiments used a fully repeated measures design. In the first half of the experimental session stimuli were presented in a pseudo-random order which avoided successive presentations of similar stimuli. This order was reversed in the second half-session, producing two judgements of step size per condition from each observer. The means of these were analysed using ANOVAs. Stimuli were varied in four respects:

(i) *stereo slant angle*, set by the physical angle of the ground or wall plane;
(ii) *texture slant angle*, set by the computer graphics;
(iii) *the physical size of the step*, set by the depth of a block of polystyrene resting on the table used in Expts 1–3, and by the separation between two sheets of polystyrene that formed the projection surfaces in Expts 4 and 5;
(iv) *the density cue between the two surfaces*, set by the computer graphics. In Expts 1–4 the density cue was consistent with the physical step size (stereo cue), in Expt 5 the density cue was consistent with the predicted step size from the model.

The number of experimental stimuli ranged between 22 and 30. Table 1 summarizes differences between the five experiments and these will be elaborated as each one is described.

*Cited significance levels are those obtained after applying, where necessary, conservative epsilon corrections for departures from covariance homogeneity assumptions (Howell, 1987). For brevity and simplicity, F values are cited only for effects of interest. In all experiments the effect of step size was significant at $P < 0.001$ or better.

Procedure

Observers read instructions prior to an initial training period during which they were aided to make subjective estimates of step sizes on a series of steps (range over all experiments: 5–25 cm). Random-line elements providing consistent texture/stereo slant cues were used for training. The observers were asked to judge the size of the step at the part of the edge shown with marker lines, see Fig. 2 and caption. Observers reported their estimates of step sizes orally. Only when they had correctly identified each training step size on two occasions, with no intervening errors during a randomly-ordered sequence of presentations, was the training deemed successful and the experiment proper started.

Time to reach the training criterion varied but was typically about 15–20 min, with the experimental session then following on and lasting about 40 min. Potential observers who failed to reach the training criterion within about 30 min were excluded from the experiment proper (about 25% of recruits were rejected in this way).

Experiment 1: steps between ground planes seen with inconsistent stereo and texture cues

The main feature of this experiment was that the stereo cue was set to 0 deg throughout, i.e. the table angle was always set parallel to the ground plane on which the observer stood [Fig. 2(a)]. Henceforth, for brevity, 0 deg table angle is referred to as "ground plane" even though the table was in fact raised about 50 cm or so above the real ground plane. Figure 2(a) also shows that the table surface slant angle, defined as the angle between the table surface and a plane perpendicular to the line of sight containing the fixation point, ranged from 40 to 60 deg (table slant angle was c. 53 deg when fixation was at the centre of the field of view).

The texture slant could take on one of three values:

$+10^\circ$ an inconsistent texture cue which made the two planes forming the step appear to slant a little "uphill" from the observer's vantage point;
0° texture consistent with stereo, hence signalling the ground plane;
-10° an inconsistent texture cue which made the two planes appear to slant somewhat "downhill".

The group mean data are shown in Fig. 3(a) for a texture comprised of a regular grid of circles of diameter 5 cm on the surface, and in Fig. 3(b) for a texture made of a sparse field of random dots (density $< 1\%$; randomly jittered from a grid of 2 cm squares on the surface). In both figures physical step size is plotted on the abscissa and mean step size judgements are plotted on the ordinate. The parameters for each curve show the texture (T) and stereo (S) cue combinations. The random-dot texture was chosen because it is known to give weak shape-from-texture effects (Stevens, 1981; Frisby & Buckley, 1992), and hence the dot conditions served as controls for which little or no texture calibration of stereo would be expected.

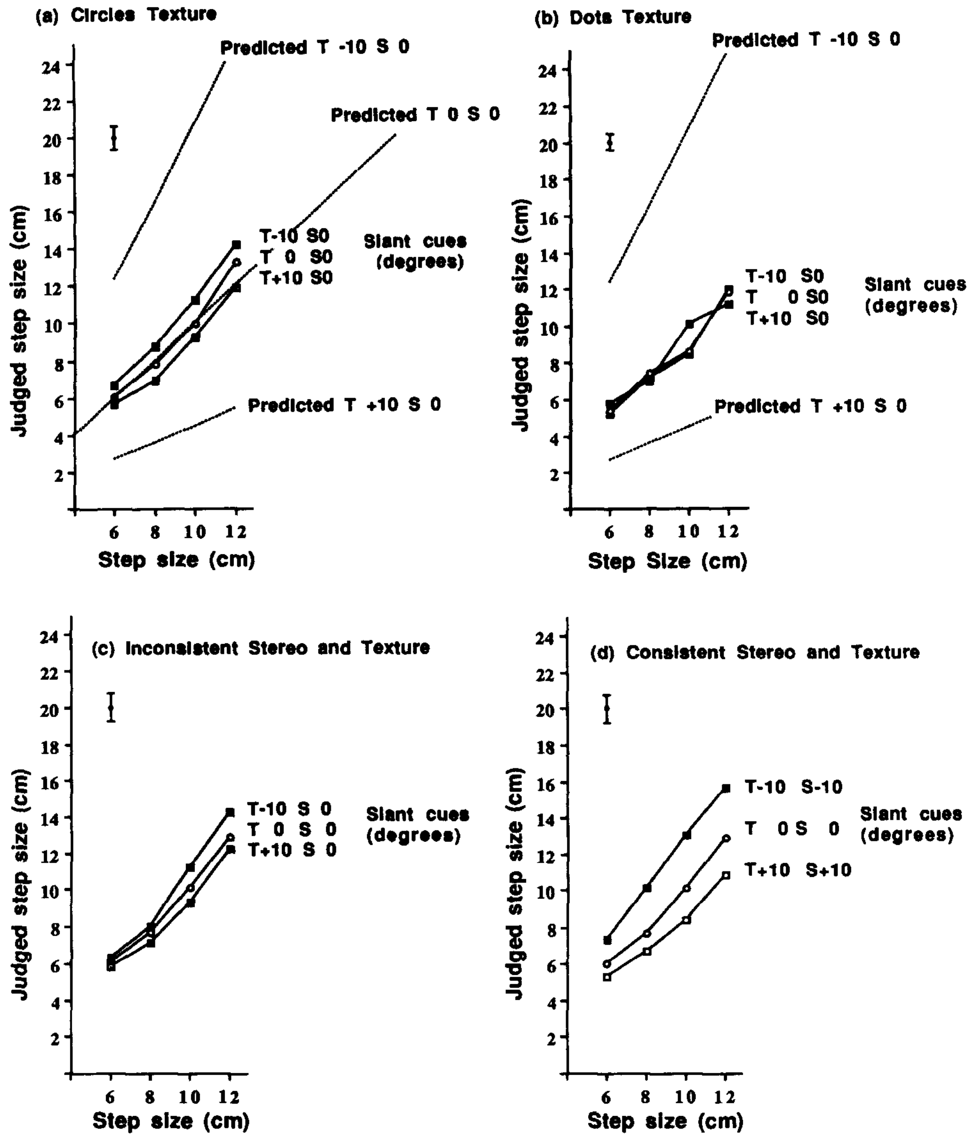

FIGURE 3. Group means: ground planes. (a) Experiment 1: effect of stereo-consistent and stereo-inconsistent texture cue (of circles) on judged step size. T, texture; S, stereo. Dotted lines indicate the predicted step sizes from the model (the Appendix gives details). (b) Experiment 1: effect of consistent and inconsistent random-dot textures on judged step size. (c) Experiment 2: effect of inconsistent circles texture on judged step size, the T = 0, S = 0 means are included for comparison. (d) Experiment 2: effect of consistent circles texture on judged step size. The error bar in the top-left corner of each graph is the mean of the 12 SEs calculated for each of the means shown from the individual observer means. This bar therefore reflects differences between observers and not the error variation used in the various ANOVAs cited in the text, as they were repeated measures designs.

The main results of interest were a significant effect of texture slant ($F_{2,10} = 18.63$, $P < 0.001$)* and a significant interaction ($F_{2,10} = 5.36$, $P < 0.05$) between the effects of texture type and texture slant angle on step size judgements. This can be seen by comparing Fig. 3(a) and Fig. 3(b): step size judgements were affected by making the texture cue inconsistent with stereo if the texture was made of circles but not if it was made of dots. The qualitative character of this result is in keeping with the hypothesis of texture calibration of stereo. Quantitatively, the effect of the circles texture was far less than predicted (the dotted lines in Fig. 3(a, b) show the predicted effects).

This quantitative undershoot, however, could easily have reflected the fact that the experimental rig provided various potential sources of stereo calibration data other than texture. For example, vertical disparities and oculomotor cues were in principle available from the table rig. If used, these would have operated against texture calibration and towards making the step sizes appear veridical. Hence Expt 1 can be said to have given some support to the hypothesis of texture calibration of stereo.

Experiment 2: steps between ground planes seen with inconsistent and consistent, stereo and circles texture cues

A problem with Expt 1 is as follows: suppose that the perceived size of a step seen between two slanted planes is determined in part by the slants of those planes. If this were so, then the observed effect of an inconsistent texture on judged step size may not have been due to texture calibration of stereo but instead to the fact that the conflicting texture cues caused the planes forming the step to appear slanted with respect to the ground plane. Although the size of this slant effect was not formally measured in the present experiments, Frisby and Buckley (1992) reported quantitative data from the same rig, albeit for planar rather than for stepped surfaces, which show that apparent slant from regular grids of circles is determined by a roughly equal weighting of stereo and texture cues. For example, they found that a mixture of 0 deg stereo and +10 deg texture produced slant judgements of about +5 deg, whereas +10 deg stereo and +10 deg texture combined to produce slant judgements of about +10 deg. Informal remarks from observers indicated that this effect also occurred in Expt 1.

Experiment 2 examined whether this effect might have influenced step size judgements. It did so by using *consistent* texture and stereo cues for uphill and downhill slants, as well as for the ground plane. Given the texture calibration of stereo hypothesis, manipulating stereo and texture while keeping them consistent should not alter step size judgements. This is because texture calibration would then lead to veridical judgements throughout. On the other hand, if step size judgements are altered as a function of the slant angles of the surfaces forming them, then these consistent cue conditions will detect that effect. The experimental procedures of Expt 2 closely followed those of Expt 1.

The main result from Expt 2 [Fig. 3(d)] was that consistent texture/stereo slant cues did indeed significantly affect judgements of step size ($F_{2,10} = 24.31$, $P < 0.001$). The consistent conditions produced a significantly greater effect than did the inconsistent cue conditions of Fig. 3(c) ($F_{1,5} = 7.35$, $P < 0.05$). This could have been due to the fact that the apparent slant in the inconsistent conditions derived from a texture/stereo cue conflict that probably produced smaller apparent slants from the ground plane than did the consistent cue conditions.

The consistent conditions of Expt 2 clearly show that step size judgements are affected by the slants of the surfaces in which they are set. This result could account for the effect observed for the inconsistent conditions, thereby casting doubt on whether that effect really does reflect texture calibration of stereo.

Experiment 3: steps between ground planes portrayed with inconsistent and consistent, stereo and dot texture cues

This experiment tested whether step judgements would be affected by consistent texture/stereo conditions using dot textures. As already noted, sparse random-dot fields create, at best, very weak texture effects in the present apparatus, making them useful as control conditions. Hence, if step size judgements were altered in consistent *dot* texture/stereo conditions then this would give added credence to the existence of an effect on step judgements arising simply from the apparent slants of the planes comprising them, unrelated to any slant-from-texture calibration effects. Experiment 3 included both consistent and inconsistent cue conditions to enable the required comparisons to be made.

The main result from Expt 3 (Fig. 4) was that consistent texture/stereo conditions altered step size judgements but the inconsistent cues did not ($F_{1,5} = 11.68$, $P < 0.05$). This therefore replicated for dot textures the pattern of results found for consistent circles in Expt 2. Moreover, there was no significant difference between the consistent cue conditions of Expt 2 (circles) and Expt 3 (dots).

By confirming that step size judgements are affected by the apparent slant angles of the surfaces comprising them, when texture and stereo cues are consistent, these results cast further doubt on the likelihood of the data from the inconsistent conditions of Expts 1 and 2 being due to texture calibration of stereo.

The next two experiments attempted to extend our tests of the texture calibration idea by using the same step size judgement task but for horizontal steps set in a "wall" plane. Each step was carefully positioned so that the observer's line of sight was always aligned with the direction of the step [Fig. 2(b)]. The advantage of this arrangement for present purposes is that the slant angle effect noted in the consistent conditions of Expts 2 and 3 may have been due to the observer's line of sight crossing

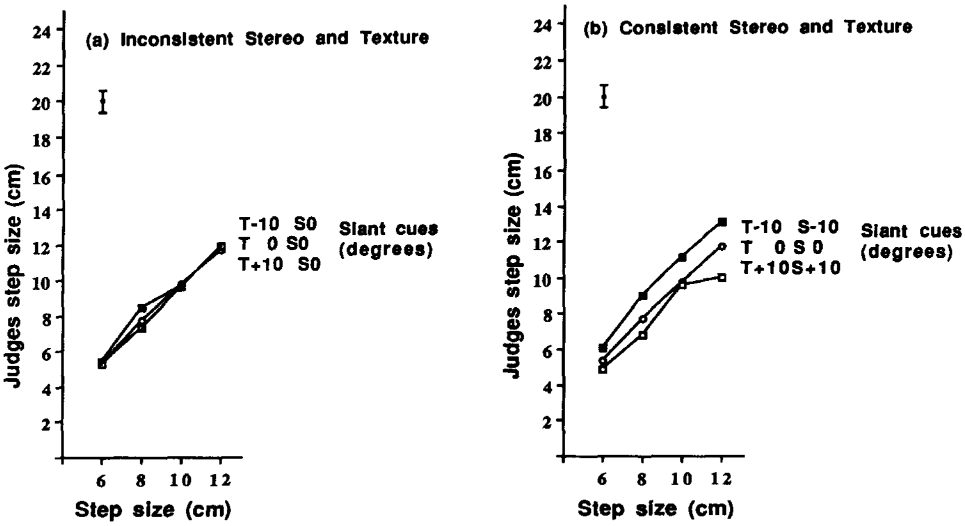

FIGURE 4. Group means from Expt 3, ground planes (see text).

the step at an angle. This may somehow have made step judgements susceptible to the slant angles of the step surfaces. Possible problems of that kind could not arise for horizontal steps viewed such that the observer's line of sight was always exactly aligned with the fall line of the step, irrespective of the slant angles of the surfaces forming the step.

Experiment 4: steps between wall planes seen with consistent and inconsistent, texture and stereo cues

The general design of Expt 4 was modelled on the earlier studies, but some changes were made to experimental parameters which derived from the wall plane apparatus. Texture and stereo slants cues were -10, -20 or -30 deg from the fronto-parallel plane, and both consistent and inconsistent conditions were included. Step sizes were 10, 15, and 20 cm. An attempt was made to strengthen the power of the texture cue by using a regular grid of squares (of 5 cm side length on the surface) instead of circles.

The group mean data from Expt 4 are shown in Fig. 5(a, b) for the consistent and inconsistent conditions respectively. As in Expt 1, inconsistent texture/stereo cues significantly affected perceived step size ($F_{2,8} = 6.51$, $P < 0.05$). This is as predicted by the texture calibration of stereo hypothesis even though the size of this effect fell well short of predictions.

However, consistent slant cues also again significantly affected perceived step size [$F_{2,8} = 12.87$, $P < 0.01$; Fig. 5(b)]. This result occurred despite the use of horizontal steps directly aligned with the observer's line of sight. It is clearly contrary to predictions from the texture calibration hypothesis and it is sufficient to explain the modest effect observed for the inconsistent conditions. Whatever the cause of this effect might be, it clearly throws doubt on interpreting the effect from the inconsistent cues as showing a texture calibration effect.

Experiment 5: steps in wall planes with consistent and inconsistent, texture and stereo cues, small (± 3 deg) conflicts between cues and density controls

One possible interpretation of the data from Expt 4 is that a much bigger effect would have arisen from the inconsistent than from the consistent conditions *if* the experimental parameters had been chosen to suit the texture calibration of stereo hypothesis more carefully. That is, there is clearly a small but reliable effect on step size judgements from alterations in the perceived slants of the surfaces forming the step, whether achieved by consistent or inconsistent texture/stereo cues. But perhaps an additional and quite separately caused effect would have been added for the inconsistent conditions arising from texture calibration of stereo *except* that this was not observed because the degree of texture/stereo inconsistency was so great that it led to a "veto" (Bülthoff & Mallot, 1988) on the use of the texture for stereo calibration purposes.

For example, projecting a texture cue of -30 deg on to a step size of 15 cm with a stereo slant of -20 deg would have required the visual system to interpret the disparity cues generated by a 15 cm step as arising from a roughly 40 cm step, given the texture calibration hypothesis. This extremely large shift might have been a wholly implausible result in view of other available cues for stereo calibration (e.g. vertical disparities etc.). Hence, perhaps

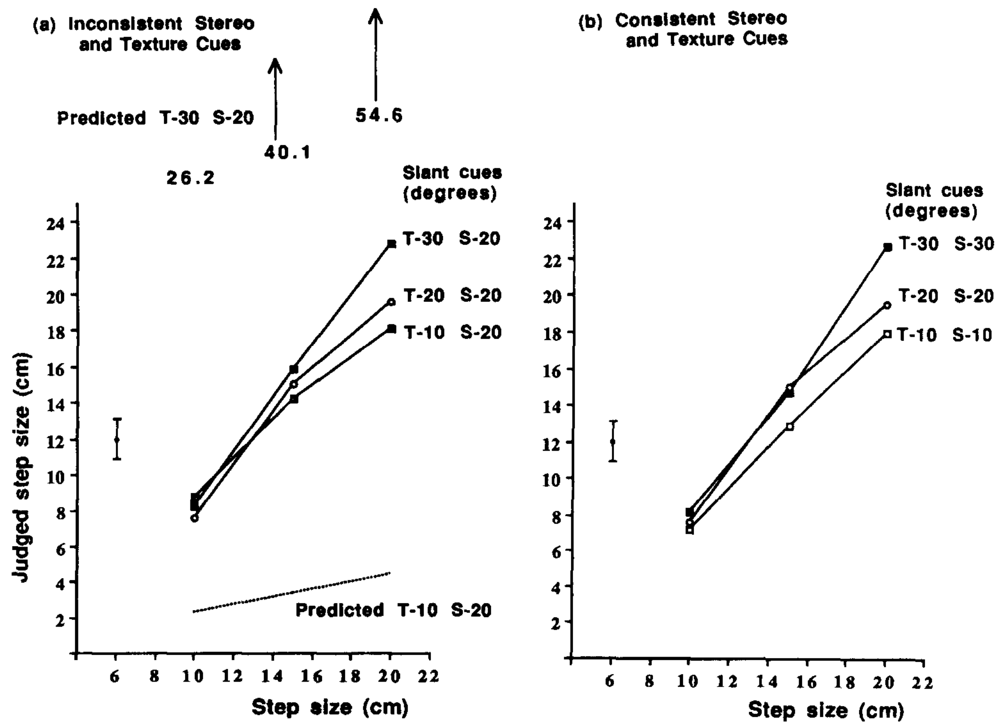

FIGURE 5. Group means from Expt 4, wall planes. Note that the predicted values for the T = -30, S = -20 stimuli in (a) are at values too great to be plotted and hence are shown as numbers.

the visual system simply suppressed its texture calibration processes in the circumstances of Expt 4.

Experiment 5 tried to circumvent this difficulty by reducing the size of the cue conflict. Because observers in the previous experiments had reported changes in step size of the order of ± 5 cm from the various conditions, texture cues were chosen that, from the model, would have created a change of that order of magnitude in an attempt to provide a more sensitive test of the texture calibration hypothesis.

Hence in Expt 5, consistent texture/stereo slant cues of -20 deg were presented for steps of 7.5, 10, 12.5, . . . , 22.5 cm. The additional critical conditions were inconsistent texture slants of -23 and -17 deg, and these were used only for a 15 cm step. Predictions from the hypothesis were that these should have been seen as approx. 20 and 10 cm steps respectively, values falling well within the range observers were expecting to see given the step range of 7.5–22.5 cm for the consistent conditions. Also included were consistent-cue control conditions for the two critical inconsistent texture slants of -23 and -17 deg. Keeping the critical inconsistent conditions down to just two was a precaution against the visual system adapting its hypothesized use of texture cues for stereo calibration when confronted by long runs of inconsistent conditions, which may have been yet another reason why large predicted effects were not observed in the earlier experiments.

A further difficulty with the previous experiments was that the step size signalled by the texture calibration processes would normally be accompanied by an appreciable difference in the density of texture elements lying on the nearer and farther surfaces of the step. Yet in fact, no such additional density change was built into the inconsistent stimuli, so that this amounted to another cue conflict operating against the outcome predicted by the texture calibration of stereo hypothesis. This was remedied in Expt 5 by changing the density across the step, in the inconsistent conditions only, so that it was congruent with the predictions (no such changes were of course required in the consistent conditions). Other aspects of the design of experiment 5 were kept the same as those of Expt 4. The results of Expt 5 are shown in Fig. 6. It reveals that the various changes made to the experimental design did not produce the predicted effects on step size judgements. Thus once again no evidence emerged in favour of the hypothesis.

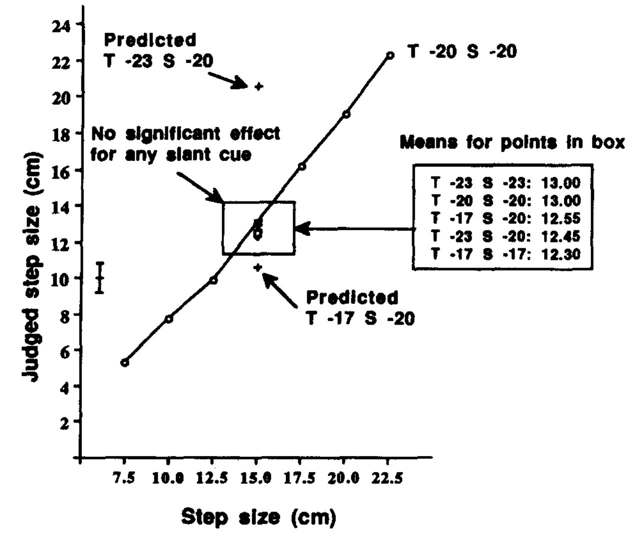

FIGURE 6. Group means from Expt 5, wall planes. Points on the line are step judgements to stimuli with stereo and texture cues consistent for a -20 deg slant ("downhill"). Points in the small central box were very similar and so the data for these conditions and the cues that generated them are shown alongside in the expanded box (see text).

DISCUSSION

The aim of the present study was to determine whether the human visual system uses slant information derived from texture to calibrate stereo. Five experiments failed to find any convincing evidence that it does so. Although step size judgements were affected by inconsistent texture manipulations, the slight observed effects also occurred for texture manipulations consistent with stereo. This indicates that whatever the cause of the observed shifts in step size judgement, their origin seems to have nothing to do with the hypothesis of texture calibration by stereo. It is of interest that this effect occurred both when the change in slant between stimuli was due to the cues being consistent and when slant percepts were changed by cue integration in the inconsistent stimuli.

Our failure to find any evidence confirming the texture/stereo calibration hypothesis does not of course mean that the visual system does not use such a method. We may have failed to discover the predicted effects for a variety of reasons. The most obvious is that in our experimental situation other data sources for stereo calibration were available (vertical disparities, oculomotor data) and that these outweighed the texture data. Perhaps it would be worthwhile testing the hypothesis in circumstances that might make those other sources less powerful, such as greater viewing distances.

It may also be that the human visual system uses texture for stereo calibration but that it does so only to confirm that the local surface planarity assumption is or is not justified, rather than actually receiving P and Q values from texture. If planarity is assumed then equation (3) for horizontal disparities can be solved if sufficient data points are available (as few as six in theory). A scheme of this type would predict that the present kind of texture slant manipulations would have no effect on step sizes because the texture slant data would not be used for stereo calibration purposes.

REFERENCES

Buckley, D. & Frisby, J. P. (1993). Interaction of stereo, texture and outline cues in the shape perception of three-dimensional ridges. *Vision Research*, *33*, 919–933.

Bülthoff, H. H. & Mallot, H. A. (1988). Interaction of depth modules: Stereo and shading. *Journal of the Optical Society of America, A*, *5*, 1749–1758.

Carpenter, R. H. S. (1988). *Movements of the eyes* (2nd edn). London: Pion Press.

Collett, T. S., Schwarz, U. & Sobel, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, *20*, 733–754.

Collewijn, H. & Erkelens, C. J. (1990). Binocular eye movements and the perception of depth. In Kowler, E. (Ed.), *Eye movements and their role in visual and cognitive process*. Amsterdam: Elsevier.

Cumming, B. G., Johnston, E. B. & Parker, A. J. (1991). Vertical disparities and perception of three-dimensional space. *Nature*, *349*, 411–413.

Enright, J. T. (1991). Exploring the third dimension with eye movements: Better than stereopsis. *Vision Research*, *31*, 1549–1562.

Foley, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*, 411–435.

Foley, J. M. (1985). Binocular distance perception: Egocentric visual task. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 133–149.

Frisby, J. P. (1984). An old illusion and a new theory of stereoscopic depth perception. *Nature*, *307*, 592–593.

Frisby, J. P. & Buckley, D. (1992). Experiments on stereo and texture cue combination in human vision using quasi-natural viewing. In Orban, G. A. & Nagel, H.-H. (Eds), *Artificial and biological vision systems*. New York: Springer.

Frisby, J. P. & Pollard, S. B. (1991). Computational issues in solving the stereo correspondence problem. In Landy, M. S. & Movshon, J. A. (Eds), *Computational models of visual processing* (pp. 331–357). Cambridge, Mass.: MIT Press.

Gårding, J. (1992). Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, *2*, 329–352.

Gårding, J., Porrill, J., Mayhew, J. E. W. & Frisby, J. P. (1995). Stereopsis, vertical disparity and relief transformations. *Vision Research*, *35*, 703–722.

Howell, D. C. (1987). *Statistical methods for psychology* (2nd edn). Boston, Mass.: Duxbury Press.

Koenderink, J. J. & van Doorn, A. J. (1976). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, *21*, 29–35.

Liu, L., Stevenson, S. B. & Schor, C. M. (1993). A polar coordinate system for describing binocular disparity. *Vision Research*, *34*, 1205–1222.

Longuet-Higgins, H. C. (1982). The role of the vertical dimension in stereoscopic vision. *Perception*, *11*, 377–386.

Mayhew, J. E. W. (1982). The interpretation of stereo-disparity information: the computation of surface orientation and depth. *Perception*, *11*, 387–403.

Mayhew, J. E. W. & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, *297*, 376–379.

Mayhew, J. E. W., Zheng, Y. & Cornell, S. (1993). The adaptive control of a four-degrees-of-freedom stereo camera head. *Philosophical Transactions of the Royal Society of London B*, *337*, 315–326.

Porrill, J., Mayhew, J. E. W. & Frisby, J. P. (1985). Cyclotorsion, conformal invariance, and induced effects in stereoscopic vision. In *Frontiers of visual science: Proceedings of the 1985 symposium* (pp. 90–108). Washington, D.C.: National Academy Press.

Porrill, J., Gårding, J., Eklundh, J.-O., Frisby, J. P., Buckley, D., Pollard, S. B., Mayhew, J. E. W. & Spivey, E. (1991). Using shape-from-texture to calibrate stereo. *Perception*, *20*, 90.

Rogers, B. J. (1991). Why the eye doesn't shape up. *Nature*, *349*, 365–366.

Rogers, B. J. (1992). The perception and representation of depth and slant in stereoscopic surfaces. In Orban, G. A. & Nagel, H.-H. (Eds), *Artificial and biological vision systems*. New York: Springer.

Rogers, B. J. & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, *361*, 253–255.

Sobel, E. C. & Collett, T. S. (1991). Does vertical disparity scale the perception of stereoscopic depth? *Proceedings of the Royal Society of London B*, *244*, 87–90.

Stenton, S. P., Frisby, J. P. & Mayhew, J. E. W. (1984). Vertical disparity pooling and the induced effect. *Nature*, *309*, 622–623.

Stevens, K. A. (1981). The information content of texture gradients. *Biological Cybernetics*, *42*, 95–105.

Turner, M. R., Gerstein, G. L. & Bajcsy, R. (1991). Underestimation of visual texture slant by human observers: A model. *Biological Cybernetics*, *65*, 215–226.

Weinshall, D. (1990). Qualitative depth from stereo with applications. *Computer Vision, Graphics and Image Processing*, *49*, 222–241.

Acknowledgements—KAW is supported by a SERC postgraduate studentship. The research was funded in part by an Esprit BRA grant within the INSIGHT consortium. Our thanks to Michael Port and Leonard Hetherington for equipment and photographic support; to Louise Hill, Eleanor Spivey and Janet Horsman for their help in conducting the experiments; and to Stephen Pollard for computer graphics advice.

APPENDIX

Using a Listing's law fixation model, given a fixation point at distance d with gaze angle g and elevation e, in the small-angle approximation the equation for horizontal disparity takes the form

$$H = \frac{I}{d}\left(\frac{Z-d}{d} + gx + \frac{1}{2}ey + x^2\right)$$

where X,Y,Z are world coordinates with respect to the cyclopean eye as origin, and $x = X/Z$, $y = Y/Z$ are image coordinates (in the body of the text z denotes depth relative to the fixation point, $z = Z - d$). This equation can be solved to recover depth Z, given horizontal disparity H. Using incorrect values d', g', e', for the calibration parameters would give an incorrectly reconstructed depth Z' with

$$H = \frac{I}{d'}\left(\frac{Z'-d'}{d'} + g'x + \frac{1}{2}e'y + x'^2\right).$$

Z and Z' are thus related by the equation

$$Z' = (d'-d) + \left(\frac{d'}{d}\right)^2 Z + \left(\frac{d'}{d}g - g'\right)X' + \frac{1}{2}\left(\frac{d'}{d}e - e'\right)Y' + \left(\frac{1}{d} - \frac{1}{d'}\right)X'^2$$

where we have used the approximation

$$X = dx, Y = dy, X' = d'x, Y' = d'y$$

to relate image coordinates to world coordinates. From left to right the terms represent

- a shift of the whole workspace by a constant distance;
- a scaling of depths relative to the fixation point;
- an added slope about a vertical axis;
- an added slope about a horizontal axis;
- an added curvature about a vertical axis (corresponding to the Veith–Muller curvature).

These corrections apply to the workspace as a whole, and in particular, to all planar elements in the workspace.We now restrict ourselves to viewing angles $g' = e' = g = e = 0$ and look at the effect on a plane

$$Z = PX + QY + R + d.$$

We find

$$Z' = \frac{d'}{d}PX' + \frac{d'}{d}QY' + d' + \left(\frac{d'}{d}\right)^2 R + \left(\frac{1}{d} - \frac{1}{d'}\right)X'^2.$$

The terms represent

- scaling of the slope about a vertical axis by a factor (d'/d);
- scaling of the slope about a horizontal axis by a factor (d'/d);
- a shift in the fixation point to d';
- scaling of distance relative to fixation point by $(d'/d)^2$;
- added Veith–Muller curvature $(1/d) - (1/d')$.

Thus if we can give a strong texture cue for an incorrect slope, Q', this is consistent with a depth calibration $d' = d(Q'/Q)$. If this incorrect calibration is accepted it will cause all the other effects outlined above. The experiments reported looked in particular for the re-calibration of relative depth $R' = R(d'/d)^2$. As pointed out by a referee, use of the incorrect calibration also predicts a change of curvature in the perceived surface; only the correct depth value results in a reconstructed planar surface. Since the slopes induced by curvature correction are small relative to the slope rescaling, and within the probable errors of a slope from texture channel, this is not an objective in principle to texture re-calibration of stereo. However it could affect the results of our experiment: assuming a special sensitivity to planarity, prior assumption of planarity would force the choice of the correct stereo calibration. This objection could be circumvented by the use of curved stimuli, in which the planar assumption would be unjustified.

A second referee commented that the planarity assumption used in deriving slope re-scaling is not satisfied by the step (two plane) stimulus. The derivation above makes it clear that the same rescaling parameter applies to all planes (indeed all surfaces) in the workspace.